

COVID-19 Case Surveillance Public Use Data Utility Summary

Users should consider the level of completeness, including suppression levels when planning their analyses and use of public datasets. Privacy protections will suppress field values to reduce reidentification risks. Completeness varies by jurisdiction (i.e., state, local, and territorial) and time period. Variables are consistently coded to the value “Unknown” when jurisdictions specify in the case data submitted to CDC that the value is unknown, the value “Missing” when jurisdictions do not provide a value, and the value “NA” when the value is suppressed as part of privacy protections.

Dataset version: 2021-04-30

Total records in dataset: 24,441,351

Quick Summary

	all_fields	quasi_fields
total_fields	12	3
total_records	24,441,351	24,441,351
total_cells	293,296,212	73,324,053
missing_fields	60,166,407	1,081,449
missing_pct	21%	1%
complete_fields	233,129,805	72,242,604
complete_pct	79%	99%
unknown_fields	19,431,207	7,761,356
unknown_pct	7%	11%
suppressed_fields	94	94
suppressed_pct	0%	0%
available_fields	213,698,504	64,481,154
available_pct	73%	88%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	7	0.0%	857,231	3.5%
sex	19	0.0%	53,752	0.2%
age_group	68	0.0%	170,466	0.7%
records_with_any_field	68	0.0%	1,021,567	4.2%